DOCUMENT RESUME

ED 353 745 EC 301 778

AUTHOR Hasselbring, Ted S.; And Others
TITLE Using Microcomputers for Assessment and Error Analysis. Monograph #23.
INSTITUTION Florida Univ., Gainesville. Dept. of Pediatrics.; Florida Univ., Gainesville. Dept. of Special Education.
SPONS AGENCY Florida State Dept. of Education, Tallahassee.
PUB DATE May 89
NOTE 47p.; A Product of the Multidisciplinary Diagnostic and Training Program.
PUB TYPE Guides - Classroom Use - Teaching Guides (For Teacher) (052)

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS *Behavior Problems; *Computer Assisted Testing; *Computer Managed Instruction; Diagnostic Teaching; Elementary Education; Elementary School Students; Error Patterns; Evaluation Methods; *Learning Problems; *Microcomputers; Scoring; Student Evaluation; Test Scoring Machines

ABSTRACT

This monograph provides an overview of computer-based assessment and error analysis in the instruction of elementary students with complex medical, learning, and/or behavioral problems. Information on generating and scoring tests using the microcomputer is offered, as are ideas for using computers in the analysis of mathematical strategies and errors of students. Examples of specific computer applications are included. Issues such as interactive assessment, curriculum-based assessment, and artificial intelligence are also considered. The systems save examiner time, reduce scoring errors, and provide teachers with information that would otherwise be difficult to obtain. More advanced systems will be able to guide the teacher through the necessary steps for the analysis of learning problems, test the student directly where appropriate, analyze student performance data, and prescribe appropriate instructional strategies. A glossary of terms concludes the monograph. (Contains 20 references.) (DB)

Using Microcomputers for Assessment and Error Analysis[1,2]

Ted S. Hasselbring, Consultant
Prisca Moore, Consultant
Jo M. Hendrickson, Assistant Professor

pediatrics • special education • psychology • speech and language •

MULTIDISCIPLINARY DIAGNOSTIC AND TRAINING PROGRAM

UNIVERSITY OF FLORIDA

Using Microcomputers for
Assessment and Error Analysis[1,2]

Ted S. Hasselbring, Consultant
Prisca Moore, Consultant
Jo M. Hendrickson, Assistant Professor

Monograph # 23

May, 1989

1 The development and publication of this monograph was supported in part by the University of Florida Multidisciplinary Diagnostic and Training Program (MDTP) funded by the Florida State Department of Education.

2 This manuscript, "Using Microcomputers for Assessment and Error Analysis," is reprinted with the permission of the authors.

THE UNIVERSITY OF FLORIDA
MULTIDISCIPLINARY DIAGNOSTIC AND TRAINING PROGRAM (MDTP)

The MDTP is administered through a joint effort by the Department of Pediatrics and the Department of Special Education at the University of Florida. The MDTP staff is composed of professionals from the fields of pediatric neurology, education, school psychology, and speech and language pathology. The MDTP has specified the elementary school student with complex medical, learning and/or behavioral problems as its primary population. Major responsibilities of the MDTP are to use all appropriate disciplines to provide diagnostic and intervention services to school systems referring students, train education and health professionals at the preservice and inservice level, and assist parents of students experiencing difficulty in school.

Co-Directors: John R. Ross, M.D.
Cecil D. Mercer, Ed.D.

Program Coordinator: Pam Walker

Program Manager: Susan K. Peterson, Ph.D.

Monograph Reviewers: Robert Gable, Old Dominion University
Lori Korinek, William and Mary University
Donna Omer, School Board of Alachua County

Multidisciplinary Diagnostic and Training Program
Box J-282 J. Hillis Miller Health Center
University of Florida
Gainesville, FL 32610
(904) 392-5874
(904) 392-6442

PREFACE

Computer technology holds great promise for improving instructional practices. This monograph provides an overview of computer-based assessment and error analysis. Information on generating and scoring tests using the microcomputer is reported. Ideas for using a computer in the analysis of math strategies and mathematical errors of students are provided. Issues such as curriculum-based assessment and artifical intelligence are discussed in a manner which should be helpful to persons without much background in comptuers. Finally, a glossary of terms concludes the monograph.

TABLE OF CONTENTS

	Page
OVERVIEW OF COMPUTER-BASED ASSESSMENT AND ERROR ANALYSIS	1
Test Generation	2
Test Scoring and Reporting	4
Interactive Assessment	11
Curriculum-Based Assessment	21
Computer-Based Monitoring and Decision Making	26
Expert Systems: The Future of Computer-Based Assessment?	28
SUMMARY	33
REFERENCES	35
GLOSSARY OF TERMS	38

Using Microcomputers for Assessment and Error Analysis

Overview of Computer-Based Assessment and Error Analysis

Numerous authors have pointed out that the assessment and analysis of error patterns, when carried out consistently and conscientiously, lead to improved student learning. Given that this is true, we must ask why student learning problems are not analyzed more frequently by classroom teachers? Bennett (1983) provides several possible reasons why error analysis is seldom carried out in the classroom. First, he suggests that too often teachers have not been trained to carry out the assessment and error analysis process. Even if teachers want to analyze student error patterns they often do not have the necessary technical skills. Second, Bennett points out that the assessment and analysis of error patterns often can be a very labor intensive process, requiring a significant amount of time and effort. When asked, most educators report that they simply do not have the time required to conduct error analysis procedures although they know error analysis is important. As a result, assessment and error analysis is a low priority activity in most classrooms.

Although this monograph should help you to develop the technical skills necessary for conducting error analyses, the problem of "too little time" is not as easily overcome. Recently, however, a number of educators have begun to exploit the power of microcomputer technology to reduce the time required to assess and analyze student learning problems. Although computer-based assessment is only in its infancy, it has already

shown potential for saving time as well as improving the accuracy of the error analysis process (Hasselbring, 1986).

The purpose of this monograph is to provide a general discussion of how microcomputers can be used in the assessment and error analysis process. When appropriate, we will provide examples of specific software programs that can be used for assessing and analyzing student performance. Throughout the monograph, we have classified computer-based assessment and error analysis software into five different categories; these include: (a) test generation, (b) test scoring and reporting, (c) interactive assessment, (d) curriculum-based assessment, and (e) expert systems. A discussion of each of these software types follows.

Test Generation

Often, the first part of conducting an analysis of student errors involves the selection or development an appropriate assessment instrument. While the process of selecting an appropriate instrument can be tedious and time consuming, selecting an assessment instrument is not nearly as burdensome as developing one. Anyone who has gone through the process of developing an assessment instrument knows that it can take many hours. Further, the problem is compounded when parallel versions of an instrument are required. Recently, however, test generation programs have been developed that can be used to remove the tedium and reduce the time it takes to create assessment instruments.

Using a test generation program, a teacher can quickly and easily create an assessment instrument that covers specific skills and objectives at a specified level of difficulty. Generally, these programs are designed to allow the teacher to build banks of test items that are classified by topic, question format, or objective and then have the computer select and print out specific assessment items based upon criteria provided by the teacher.

One such example of a test generation program is called Exam. Developed by the Brownstone Research Group, Exam can be used to create assessment instruments as well as keep computerized records of student performances on these tests. Before generating a test, the teacher must first use Exam to develop a bank of test items covering the topic or skills to be assessed. The test items can be written in a variety of formats which include: multiple choice, true/false, matching, or essay. In addition to categorizing assessment items by type, Exam allows the teacher to categorize items by topic or level of difficulty. Once the test items are developed they can be stored on disk for future use. The stored item bank can be edited at any time by adding items, deleting items, or by changing the wording of any question.

Following the creation of an item bank, Exam can be used to select specific items from the bank and print an assessment instrument. The teacher enters the criteria to be used for selecting items from the bank, such as difficulty level and type

of item, and the computer does the rest. A variety of instruments can be created using the same item bank by altering the type of questions, the order of questions, or the difficulty level. This gives the teacher an opportunity to individualize the evaluation of students by creating and administering different versions of an assessment instrument that covers the same information.

By allowing the computer to assist in the development of assessment instruments, teachers can quickly and easily generate a series of tests. If these instruments are carefully sequenced with regard to item difficulty, the teacher is able to determine at what point a child's skills break down or in what specific area. Not only is this information useful for developing remediation programs but this information is also useful for documenting when and what skills have been mastered.

Test Scoring and Reporting

Test generation programs like the one described above are useful in as much as they allow teachers to create tests which serve as diagnostic tools. However, after generating the test and giving it to the student the teacher must still score the student responses and analyze the results. Currently, there are computer programs that can score and summarize student responses on an assessment. The earliest of these programs were developed primarily for school psychologists to allow them to quickly score and summarize standardized tests such as the WISC-R, WAIS-R, PIAT, and Woodcock-Johnson Psycho-Educational Battery. These

programs allowed psychologists to be more productive since they no longer had to spend hours scoring and summarizing the tests manually. The computer was able to do much of this for them.

When using a scoring program, the test is administered in its traditional form which is usually a paper and pencil format, or in some cases orally, which means the examiner must record the student's response. In neither case does the student interact with the computer, only the examiner. The examiner must enter a student's scores from the administration of the test into the computer. This is generally done in one of two ways, either by typing the information through the keyboard or by having the computer read the information from an optical scanning sheet. After entering the student's assessment information, the computer then summarizes and prints out the results in report form. The primary advantage of this type of program is that it saves time by freeing the examiner from such clerical tasks as adding raw scores, looking through conversion tables, and in some cases writing reports.

Recently, a number of scoring programs have been developed specifically for classroom teachers. One example is the PRO-SCORE Systems by Pro-Ed. In each of the PRO-SCORE programs, the examiner first administers the test to the student in the traditional manner and then enters the student's responses into the program. The computer then generates a multi-page report that includes: raw scores, standard scores, percentiles, and descriptors for each subtest as well as a cognitive aptitude

score. Although these programs do not provide detailed error analyses, the information that is provided can be useful as part of the overall assessment process.

One program designed specifically to score, analyze, and report student spelling errors is the Computerized Test of Spelling Errors (CTSE) (Hasselbring, 1984). Using the CTSE, the teacher types in the student's spellings to a list of 40 preselected words, and the program analyzes the spellings and generates a report of the types of errors exhibited by the student. The report that is generated from the program is divided into four sections. The first section provides demographic data on the student and the testing session. The second section simply summarizes the number of words that were spelled correctly and the number spelled incorrectly. Section three of the report lists the student's spelling of all words, both correct and incorrect. Lastly, section four provides an analysis of the types of words spelled, types of errors, and error tendencies exhibited by the student. A sample report from the CTSE is shown in Figure 1.

Figure 1. Sample Report on the Computerized Test of Spelling Errors

I. Demographic Data

Student's Name: Ben H.	Date of Examination: 04/18/88
Birthdate: 06/18/80	Chronological Age: 7 - 10
Grade: 2	School: Oak Hill
Examiner: TSH	Place of Examination: Oak Hill

Elapsed Examination Time: 00:30:58

II. Summary

Number of Words Correct: 20	Percent of Words Correct: 50%
Number of Words Incorrect: 20	Percent of Words Incorrect: 50%

III. Student's Responses

Words Spelled Correctly	Words Spelled Incorrectly
ARROW	ANKEL (ANKLE)
ATE	FLYS (FLIES)
BULL	MILLON (MELON)
BAKE	ASELP (ASLEEP)
AWHILE	GUINT (GIANT)
BORN	LATTER (LATER)
DARK	JUMPPING (JUMPING)
HER	DURKED (DROPPED)
CUTTING	POW (PAW)
LAUGHING	FEILD (FIELD)
AM	THOW (THOUGH)
DRUM	HAMER (HAMMER)

LOOKED	CIRKER (CRACKER)
GIRL	BABYS (BABIES)
HOPED	ALOON (ALONE)
PONY	PARTYES (PARTIES)
BROOM	OBAY (OBEY)
MARCH	FATHER (FARTHER)
STREET	PEICE (PIECE)
BOXES	AME (AIM)

Computerized Test of Spelling Errors Evaluation Summary

IV. Diagnostic Error Analysis

		Number Incorrect	Percent Incorrect
A.	Word Types		
1.	Regular	1 out of 6	16
2.	Predictable	12 out of 20	60
3.	Irregular	7 out of 14	50
B.	Error Types		
	Vowels		
1.	Schwa vowel sound	1 out of 5	20
2.	Short vowel sound	3 out of 10	30
3.	Long vowel sound	0 out of 11	0
4.	Digraphs and diphthongs	6 out of 11	54
5.	R-controlled vowels	1 out of 10	10
	Consonants		
6.	Initial position	0 out of 24	0
7.	Medial position	1 out of 23	4
8.	Final position	0 out of 12	0
9.	Blends and digraphs	4 out of 13	30

Generalizations and Patterns

10. Suffixes	2 out of 11	18
11. Affixation rules	4 out of 5	80
12. Orthographic patterns	3 out of 6	50
13. Final E rule	1 out of 5	20

C. Error Tendencies

14. Omissions	Number of words with omissions = 8
15. Insertions	Number of words with insertions = 3
16. Substitutions	Number of words with substitutions = 14
17. Order Errors	Number of words with order errors = 5
	Total number of order errors = 5

In summary, test scoring programs like the ones described can be useful in that they reduce the amount of time associated with the scoring and error analysis process. However, on the negative side, these programs still require that the teacher enter the assessment information into the computer. In the next section, we will focus on programs that allow testing to be done directly on the computer.

Interactive Assessment

In many respects the types of programs discussed thus far do not take full advantage of the computer. An area of computer-based assessment that is gaining in popularity and does take advantage of many of the powerful characteristics of the computer is interactive assessment. Interactive assessment differs from the programs just described in that the computer plays the role of the examiner and carries out and analyzes the assessment data. The advantages of interactive assessment are obvious. For one, huge savings in examiner time can be accrued. This is especially important if teachers want to monitor student progress on a regular basis or if they want to monitor the progress of several students. Also, in some cases, examiner bias, administration errors, scoring errors, and invalid or erroneous analyses and interpretations can be more tightly controlled or eliminated. However, on the negative side, with interactive assessment the teacher is removed from the assessment process. In some cases this can create a black-box phenomenon where it may be unclear as to how the computer came up with the analysis.

Nevertheless, we believe that interactive assessment programs can with very little effort provide extremely useful data to the classroom teacher. As these programs become more sophisticated and better designed, they will become an important part of the assessment and error analysis process.

For the purpose of describing this new and exciting form of assessment, we have selected two interactive assessment programs in the area of mathematics. Although these programs are math oriented, from the descriptions you will see that they provide very different data for the classroom teacher; they were selected for this reason.

Chronometric Analysis of Math Strategies

Today, many teachers and parents are content when students with learning handicaps can compute answers to basic math facts using counting strategies (i.e., fingers and number lines) or electronic calculators. However, research by Resnick (1983) suggests that these procedures can interfere with the learning of higher level math skills such as multiple digit addition and subtraction, long division, and fractions. Most cognitive scientists today believe that as basic math skills become more highly practiced, their execution requires less cognitive processing capacity, or attention, and the student becomes fluent. Since all people have a limited capacity for information processing, not having to use part of this limited capacity for performing basic skills means that there is more capacity left for executing higher-order processes. Thus, it appears that the

ability to succeed in higher-order processes is related directly to the efficiency with which these lower level skills are executed.

Recent studies have shown that by isolating non-fluent math facts and providing individualized computer-mediated training and practice on these facts that even students with learning problems can learn to retrieve the facts from memory (Hasselbring, Goin, & Bransford, 1987). A key component of this training relies on the ability to identify each student's repertoire of fluent and non-fluent facts.

The best way to determine if a fact is fluent or not is to record how long it takes a student to answer a problem. For example, most adults answer basic facts under .6 seconds while some students may take up to 10 seconds or more because of inefficient strategies they are using. Until recently, it has been virtually impossible to record accurately the response latencies for individual math facts outside of a laboratory setting. Further, it is impossible to tell from traditional paper and pencil forms of assessment which facts have been memorized and which have not. Thus, teachers have been denied this valuable source of assessment data. However, by providing teachers with a tool that allows them to easily and reliably assess and monitor response latencies they will be better able to provide students with the fluency training that they need.

To make the classroom assessment of fluency feasible, an interactive computer program called CAMS (Hasselbring & Goin,

1985) was developed. CAMS is an acronym for the Chronometric Analysis of Math Strategies. CAMS is designed to record and analyze student response latencies for all basic facts in the four operations. Basic addition facts are defined as all facts from 0+0 to 12+12. Basic subtraction facts are defined as all facts from 0-0 to 24-12 with the subtrahend always being between 0 and 12. Multiplication includes 0x0 to 12x12 and division 0/0 to 144/12 with the divisor always being between 0 and 12.

Prior to beginning the actual assessment, the student is given several days of keyboarding practice on numbers. During the practice periods, the computer records the response latency for each number between 0 and 24. The response times for each set of numbers are used in the data analysis to factor out motor response time from the actual computation time. In other words, the time required to find and press the number key is subtracted from the total time required to solve the problem.

The testing of the basic facts is done interactively, that is, the student takes the assessment on the computer. The assessments for each of the four operations are given independently. CAMS presents the problems to the student on the video display unit of the computer. Once the assessment begins, the student is presented with a discrete set of facts one at a time. The student responds by typing the answer to the problem using the number keys on the top row of the computer keyboard and then presses the space bar to have a new problem presented. The student's response is timed from the presentation of the problem

to the pressing of the number key. If a two or three digit answer is required, the timing stops with the pressing of the second or third number key. Student response latencies are recorded to the nearest one hundredth of a second.

CAMS provides the student with a tailored assessment. That is, CAMS decides which problems the student receives based upon responses to past problems. So, for example, if a student has had difficulty with problems from the six and seven tables in multiplication, CAMS will not continue to give the student problems from these tables. CAMS attempts to predict the problems that the student is able to answer and avoid problems that are too difficult. Thus CAMS avoids being overly frustrating for the student yet provides extremely rich data with regard to the student's level of fluency. Upon completion of the assessment, CAMS analyzes the data and provides a three-part printed summary which includes a: (a) chronometric analysis, (b) descriptive analysis, and (c) fluency matrix.

The first section of the CAMS Report is a chronometric analysis profile. By plotting a student's latency data a visual profile of the student's strategy for solving the basic facts becomes apparent. The second section of the CAMS Report provides descriptive data on the student's level of fluency. This analysis provides data on the number of fluent facts, total number of problems presented, the accuracy of responses, and the number of correct, incorrect

responses. The final section of the report is a matrix. Shaded cells of the matrix indicate fluent facts. Unshaded cells indicate non-fluent facts.

In sum, CAMS provides the teacher with extremely detailed information concerning a student's level of mathematical fluency at a very low cost in terms of teacher time and effort. With CAMS, the true power of the computer is exploited. The assessment is presented interactively on the computer, the test items are tailored to the student based upon decisions that the computer makes during the assessment process, scoring is done totally by the computer, and reports are generated that allow the teacher of quickly analyze the students level of performance.

Error Analysis in Mathematical Computation

Although the importance of fluency in basic facts cannot be stressed too heavily, teachers are cautioned against the conclusion that if students have developed fluency in the basic math skills that they naturally will become better at higher-order computations. On the contrary, the development of the effortless recall of basic math facts simply permits the learner to allocate a greater proportion of attention resources to the higher-level processes. Systematic instruction in the higher-order skills is necessary if these students are to master these skills. Thus, an important part of this instruction is the identification and analysis of faulty algorithms that students may develop.

Normally, faulty algorithms are analyzed by having a student

complete a worksheet containing a variety of problems and for all incorrect answers try to determine why the student missed the problem. Often, the reasons are quite apparent, but at other times trying to determine where the student went wrong is time-consuming and difficult.

One solution is to have the computer analyze student responses and determine the cause of any errors. One set of programs that provide this type of analysis in addition, subtraction, multiplication, and division is the Math Assistant series by Scholastic Inc. The Math Assistant series has several of the characteristics of computer-based assessment programs described throughout this chapter. For example, Math Assistant can be used to create diagnostic tests at any level of difficulty by entering the problems manually or having the computer generate them automatically; students can take tests at the computer or on printed tests at their desks; when tests are taken on the computer, student responses are saved automatically and can be analyzed immediately by the computer, but if students take the test on paper the students' answers must be typed into the computer manually for error analysis.

Error reports can be printed after student test answers have been entered into the computer either by test-taking on the computer or by entering them manually. Either individual or group error reports are available. An example of an individual report is shown in Figure 2.

Figure 2. Sample Individual Report on <u>Math Assistant</u> Individual Reports

Individual Report A	Subtraction

This is a record of student errors on all tests. The error number is listed for every incorrect answer. 'C' shows a correct answer. '*' marks when no answer was given. '#' means that no answer was required (because the test contained fewer than 20 problems).

PROBLEM NUMBER

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20

NAME: JAMES ROSE TEST: LEVEL ONE/ SUB SCORE: 50%

16 1 10 C C C 1 12 C C # # # # # # # # #
#

Individual Report B — Subtraction

This report shows the total number of times (in parentheses) each error was made on all tests.

Error	1	JAMES ROSE	(3)
Error	3	JAMES ROSE	(1)
Error	10	JAMES ROSE	(1)
Error	12	JAMES ROSE	(1)
Error	16	JAMES ROSE	(1)

Individual Reports provide a summary of the number and percent of problems answered correctly, which problems were answered correctly, and the type of error made on problems answered incorrectly. Unfortunately, in order to see the problems that the student missed the test must be printed out. Also, the error type is identified only by a number. In order to identify the error type one must go to the manual for a descriptor. For example, Error 1 is represents a regrouping problem while Error 10 represents an error on basic facts. Further, when multiple errors occur on a single problem, only a single occurrence shows on Report A. Double errors are shown only on Report B.

Group Reports are virtually identical to Individual Reports except that error data are reported for multiple students. An example of a Group Report is shown in Figure 3. As can be seen, type 10 errors were made by four of the seven students. With this information, teachers can prioritize their instructional objectives for group and individual teaching formats.

In sum, Math Assistant provides a useful error analysis of student computation errors in the four operations. Perhaps Math Assistant and other similar programs can be used most productively by having students take tests on a regular basis so that errors can be monitored on over time to determine if the instructional program is effective. Since it takes only a few minutes to prepare a test using Math Assistant it provides an easy way to regularly monitor student performance. Ideally,

these performance data should be used in conjunction with curriculum-based assessment software described in the next section. In this way, student progress can be monitored in a systematic manner.

Curriculum-Based Assessment

In the past two decades, special education teachers have been trained to collect daily performance data to monitor the progress of their students. For example, consider an instructional program designed to teach a student place-value subtraction. Each day the teacher instructs on the rules of subtraction and then probes the student on 40 problems which test the concepts of place-value subtraction. On each probe the teacher records the amount of time it takes the student to complete the 40 problems and records the number of correct and incorrect responses.

The purpose of collecting daily data is to enable the teacher to determine whether the instructional program is working as planned. Specific techniques have been developed to enable teachers to use this type of classroom data to determine when an instructional strategy should be changed. Haring, Liberty, and White (1980) developed a set of guidelines, called data-based decision rules, to help teachers determine not only when instructional strategy should be changed, but also what kind of change is likely to produce favorable results for a particular student at a particular time. The decision rules help the teacher choose the strategy that has the highest probability of success.

Figure 3. Sample Group Report on Math Assistant

Group Report A	Subtraction

This is a record of student errors on all tests. The error number is listed for every incorrect answer. 'C' shows a correct answer. '*' marks when no answer was given. '#' means that no answer was required (because the test contained fewer than 20 problems).

PROBLEM NUMBER

NAME: BENJAMIN LEE TEST: LEVEL ONE/ SUB SCORE: 100%

C C C C C C C C C C # # # # # # # # #
#

NAME: JAMES ROSE TEST: LEVEL ONE/ SUB SCORE: 50%

16 1 10 C C C 1 12 C C # # # # # # # # #
#

NAME: CATHY PARSONS TEST: LEVEL ONE/ SUB SCORE: 70%

2 10 15 C C C C C C C # # # # # # # # # #

NAME: BENJAMIN LEE TEST: LEVEL TWO/ SUB SCORE: 90%

10 C C C C C C C C C # # # # # # # # # #

NAME: CATHY PARSONS TEST: LEVEL TWO/ SUB SCORE: 60%

2 15 13 10 C C C C C C # # # # # # # # # #

NAME: SAMUAL JACKSON TEST: LEVEL THREE/ SUB SCORE: 90%

10 C C C C C C C C C # # # # # # # # # #

NAME: BENJAMIN LEE TEST: LEVEL THREE/ SUB SCORE: 70%

1 10 C C C C C C C 10 # # # # # # # # # #

Group Report B	Subtraction

This report shows the total number of times (in parentheses) each error was made on all tests.

Error 1	JAMES ROSE	(3)
	BENJAMIN LEE	(1)
Error 2	CATHY PARSONS	(2)
Error 3	JAMES ROSE	(1)
Error 10	JAMES ROSE	(1)
	CATHY PARSONS	(2)
	BENJAMIN LEE	(3)
	SAMUAL JACKSON	(1)
Error 12	JAMES ROSE	(1)
Error 13	CATHY PARSONS	(1)
Error 15	CATHY PARSONS	(2)
Error 16	JAMES ROSE	(1)

In order to use these decision rules in the conventional manner, the teacher plots the data on semi-logarithmic graph paper. The initial three days of data are plotted as a baseline, and an "aim star" is drawn at the intersection of the desired level of performance and the target date for achieving that level of performance. A "minimum celeration line" is then drawn from the midpoint of the baseline data to the aim star. This line indicates the minimum level of acceleration or deceleration necessary to achieve the criterion level of performance by the target date.

As the teacher continues to conduct the instructional program, she collects and charts data. Three consecutive days of data falling below the 'celeration line indicate that the student is not learning satisfactorily and that a change should be made in the instructional strategy. In addition, a "line of progress" is drawn between the median of the most recent three days of data and the median of the three previous days of data to determine the trend of the student's performance. If the student is not progressing and/or has fallen below the `celeration line', and additional flow chart is used to determine what type of change in the instructional strategy is most likely to be successful.

Research on the effectiveness of curriculum-based assessment has indicated that this methodology is quite promising for improving student achievement. For example, Fuchs and Fuchs (1986) analyzed 21 research studies that evaluated CBA procedures. The results of this analysis indicated the use of

CBA procedures significantly increased the academic achievement of students whose teachers used these procedures. From these findings, we can assume that when a student's instructional programs are monitored using CBA procedures they will achieve much more than students whose programs are not monitored using a CBA approach.

Fuchs and Fuchs conclude that the use of CBA procedures increases student academic achievement and that the greatest gains can be expected when teachers use specific data-based rules and graph performance data for making instructional decisions. Despite the apparent effectiveness of CBA procedures, all indications are that teachers are reluctant to employ them. In a national survey of LD teachers, Wesson, King, and Deno (1984) found that, although teachers believe that CBA procedures are effective, they do not use the methodology because it is too time consuming.

Computer-Based Monitoring and Decision Making

In an attempt to make curriculum-based assessment less time consuming and easier for teachers to implement, a number of developers have proposed the use of microcomputers for implementing CBA procedures (Fuchs, Fuchs, Hamlett, & Hasselbring, 1987; Fuchs, Deno, & Mirkin, 1984; Hasselbring & Hamlett, 1984; West, Young, & Johnson, 1984). Basically, these monitoring programs have been designed to assist teachers in storing, graphing, and analyzing student performance data. One

such computer program that has been used successfully in special education is AIMSTAR.

AIMSTAR is an integrated set of computer programs that are designed assist teachers in storing, graphing, and analyzing student performance data. To use AIMSTAR, the teacher creates a student data file. Descriptive information about the student's instructional program, the program objectives, and teaching procedures are included. Following each teaching session, the teacher enters student performance data into the computer. For example, the number of correct and incorrect responses exhibited by the student and the amount of time required for the student to complete the trialsis entered. AIMSTAR then stores this information and allows the teacher to graph the student's data, apply data-based decision rules, and produce a printout giving the status of the student's instructional program with recommended changes when appropriate. The advantage of this type of analysis and report is that it is immediately obvious when a student is having difficulty with a skill and when an instructional strategy is or is not working.

It should be emphasized that monitoring and decision making programs, such as AIMSTAR, do not eliminate the need for teacher intuition and judgment in planning instruction. Rather, these programs supplement teacher judgment by providing additional empirical data and analytic procedures. Teachers using this technology are able to respond more flexibly and effectively to changing student needs and to produce greater student growth.

Expert Systems: The Future of Computer-Based Assessment?

Thus far, we have discussed assessment and error analysis programs that are used primarily to reduce the tedium associated with the analysis of student learning problems. For the most part, however, these programs have in no way provided the same intelligent insight into the assessment and error analysis process that human diagnosticians can provide. But what does the future of computer-based assessment and error analysis hold? Consider the following scenario: A teacher in a rural area is perplexed by a learning problem exhibited by one of his students. The school is too small to have the regular services of a school psychologist or diagnostician. However, with this child, the teacher feels that he needs an expert's advice to help with an analysis of the student's problem and to recommend an effective instructional program. So he goes to the computer at his desk and enters into a dialogue with it. Through the program he is using he has access to the knowledge, judgment, and intuition of the country's best educational diagnosticians. The program queries the teacher concerning the student's problems; requesting information from the teacher that will help the system come up with an analysis and prescription that will have a high probability for success. In short, the computer serves the role of an expert consultant that the teacher can call upon at any time and discuss problems that students are having in the classroom.

Unfortunately, the gap between the scenario above and reality is large. Nevertheless, both computer and cognitive scientists are making great strides in producing intelligent consultants called expert systems. Expert systems have evolved over the past 25 years from the field of artificial intelligence and can be defined as computer programs capable of reaching a level of performance comparable to that of a human expert in some specialized problem domain (Nau, 1983). Expert systems are unlike conventional application programs in that they are the first systems designed to help humans solve complex problems in a common sense way. These systems use the methods and information acquired and developed by a human expert to solve problems, make predictions, suggest possible treatments, and offer advice that is as accurate as its human counterpart.

While expert systems in the field of education are only in the early phases of development, expert systems are being used more widely in other fields. Some of the expert systems currently in use include: Caduceus, which helps doctors diagnose medical problems; CATCH, which can scan 250,000 photographs to assist New York City police in identifying criminal suspects; and Prospector, which sifts geological data to predict the location of mineral deposits. The practical applications for expert systems abound. Whenever human experts are in great demand and short supply, a computer-based consultant can help to amplify and disseminate the needed expertise.

Expert systems are designed to perform the same role as the human expert consultant, that is, provide advice in situations where highly specific knowledge and experience is needed. Ideally, an expert system should provide the user with the same information and play the same role as a human expert when placed in the same situation. For example, a medically-oriented expert system might guide an inexperienced intern by asking relevant questions about the case. The intern would respond by typing the answers to these questions. The intern continues this dialogue with the computer until the expert system has sufficient data on which to make a diagnosis. At this point, the intern either accepts the diagnosis or simply uses the computer's diagnosis as another piece of data since these systems are intended to complement, not replace, a physician's judgment and intuition.

The development of an expert system is an extremely laborious process. It requires that a human expert (or group of experts) in some domain is willing to spend hundreds of hours explaining to a "knowledge engineer" how s/he solves particular problems. It is the responsibility of the "knowledge engineer" to develop a knowledge base and set of decision rules that represent the thinking process of the expert.

Current expert systems technology seems best suited to diagnosis or classification problems whose solutions depend primarily on the possession of a large amount of specialized, factual, and empirical knowledge (Duda & Shortliffe, 1983). Thus, it is only logical that expert systems be developed in

education for assessing and diagnosing error patterns and learning problems. Although Math Assistant, discussed earlier in this chapter, was designed to provide the teacher with expert-like information, by definition, Math Assistant is not a true expert system.

To date, several expert systems have been developed that are designed to assist in the diagnosis and analysis of student learning problems. One such prototypic system developed by Colbourn and McLeod (1983) assists teachers in the diagnosis and analysis of reading problems. This system guides the teacher through reading diagnosis from the initial suspicion that a reading problem exists to the point at which sufficient information has been gathered to plan an appropriate remedial program.

With this system, a dialogue is conducted between the user and the computer, with the system posing questions or making appropriate suggestions. If desired information is not available the system provides the diagnostician the option of stopping the dialogue in order to obtain the needed data. In some cases where it is impossible to obtain the desired information the system is capable of handling incomplete data. However, in the case when the system must have further input in order to continue the diagnosis, the system reiterates what data are required and then terminates the session.

This expert system does not test the student directly, nor does it manage the testing activities. Instead, the teacher or

diagnostician performs the tasks suggested by the system (such as administering a specific test) and enters this information into the system. After these new data have been entered, the system analyzes this information and proposes the next step in the assessment process. When a sufficient amount of information has been gathered and entered, the system provides a report of its diagnostic findings. The teacher can then plan a remedial program based on these results. An obvious extension of this system is to have it prescribe appropriate remedial strategies and instructional techniques based upon the diagnostic findings.

The performance of this expert diagnostic system has been evaluated by comparing it against human diagnosticians. When subjected to a number of test cases it was found that the expert system's diagnostic reports were consistently good. In contrast, the diagnostic reports prepared by the human experts varied dramatically in terms of style, format, readability, relevance, and accuracy. Of course, some of the reports of the human diagnosticians were judged better than the reports of the expert system while many others were judged as inferior.

Other examples of expert systems capable of diagnosing and analyzing student learning problems exist and others are currently under development. DEBUGGY, one of the earliest expert systems designed for education (Burton, 1982), is used for diagnosing student errors or "bugs" in the domain of place-value subtraction. The system is based on a model where student errors are not seen as random but rather as predictable bugs in an

algorithm. After analyzing student errors in subtraction the expert system attempts to determine the student's "buggy" rules for solving the problem. DEBUGGY does this by hypothesizing the student's bugs; then it attempts to predict not only whether the student will get a similar problem incorrect but also what the exact answer will be. This system has been used quite successfully for diagnosing faulty algorithms commonly used by students when solving subtraction problems.

It appears that the development of expert systems for assessing learning problems shows great promise. However, we must view the emergence of expert systems with mixed emotion. If expert systems are developed with pedagogical soundness then these systems will no doubt benefit teachers and students. On the other hand, it is very likely that a number of systems that are pedagogically unsound will find their way to the educational marketplace. Accordingly, we must be cautious and evaluate very carefully any computer program that proposes to diagnose and analyze student learning problems.

<u>Summary</u>

Computer-based assessment and error analysis, while only beginning to be used in the schools, appears to offer great promise for overcoming many of the problems associated with the assessment of student's learning difficulties. When programmed appropriately, microcomputers can remove much of the tedium associated with the administration and scoring of diagnostic instruments. With assessment programs becoming more

sophisticated, the computer is playing a larger role in the analysis of learning problems. For example, in the case of the CAMS program described earlier in this chapter, the computer presents the student with appropriate assessment items, monitors student responses, scores, summarizes, and prints out a report on the student's performance. CAMS and other analysis programs are relatively simple in comparison to the expert systems being developed currently, nevertheless, they have been shown to be extremely successful for saving examiner time, reducing scoring errors, and providing teachers with information that would be difficult to obtain without the use of a computer.

The assessment and analysis programs currently in use are only precursors to the more elaborate and powerful intelligent systems that will be available in the next five to ten years. These systems will be able to guide the teacher through the necessary steps for the analysis of learning problems, test the student directly where appropriate, analyze student performance data, and prescribe appropriate instructional strategies for remediating the student's problems. Although it is unlikely that the use of computers will eliminate all of the difficulties involved in assessing and analyzing learning problems, existing research suggests that the process can be enhanced through the responsible use of this technology.

References

Bennett, R.E. (1983). Research and evaluation priorities for special education assessment. Exceptional Children, 50(2), 110-117.

Brownstone Research Group. (1987). Exam [computer program]. Denver, CO: Brownstone Research Group.

Burton, R. (1982). Diagnosing bugs in simple procedural skill. In D. Sleeman & J. Brown (Eds.), Intelligent tutoring systems (pp. 157-183). New York: Academic Press.

Colbourn, M., & McLeod, J. (1983). Computer-guided educational diagnosis: A prototype expert system. Journal of Special Education Technology, 6, 30-39.

Duda, R., & Shortliff, E. (1983). Expert systems research. Science, 220, 261-268.

Fuchs, L., Deno, S., & Mirkin, P. (1984). The effects of frequent curriculum-based measurement and evaluation on pedagogy, student achievement, and student awareness of learning. American Educational Research Journal, 21, 449-460.

Fuchs, L., & Fuchs, D. (1986). Effects of systematic formative evaluation: A meta-analysis. Exceptional Children, 53(3), 199-208.

Fuchs, L.S., Fuchs, D., Hamlett, C.L., & Hasselbring, T.S. (1987). Using computers with curriculum-based monitoring: Effects on teacher efficiency and satisfaction. Journal of Special Education Technology, 8(4), 14-27.

Haring, N., White, K., & White, O. (1980). Rules for data-based strategy decisions in instructional programs. In W. Sailor, B. Wilcox, & L. Brown (Eds.), Methods for instruction for severely handicapped students (pp. 159-192). Baltimore, MD: Paul H. Brookes.

Hasselbring, T.S. (1984). Computerized test of spelling errors [computer program]. Nashville, TN: Expert Systems Software, Inc.

Hasselbring, T.S. (1986). Toward the development of expert assessment systems. Special Services in the Schools, 2(2/3), 43-56.

Hasselbring, T.S., Goin, L.I. (1985). Chronometric analysis of math strategies [computer program]. Nashville, TN: Expert Systems Software, Inc.

Hasselbring, T.S., Goin, L.I., & Bransford, J.D. (1987). Effective mathematics instruction: Developing automaticity. Teaching Exceptional Children, 19(3), 30-33.

Hasselbring, T.S., & Hamlett, C.L. (1984). Planning and managing instruction: Computer-based decision-making. Teaching Exceptional Children, 16(4), 248-252.

Nau, D. (1983). Expert computer systems. IEEE Computer, 16, 63-85.

Pro-Ed. (1987). PRO-SCORE systems [computer program]. Austin, TX: Pro-Ed.

Scholastic Inc. (1985). Math assistant [computer program]. New York: Scholastic Inc.

Resnick, L. B. (1983). A development theory of number understanding. In Herbert P. Ginsburg (Ed.), _The development of mathematical thinking_ (pp. 109-151). New York: Academic Press.

Wesson, C., King, R., & Deno, S. (1984). Direct and frequent measurement: If it's so good for us, why don't we use it? _Learning Disability Quarterly_, _7_(1), 45-48.

West, R., Young, R., & Johnson, J. (1984). Using microcomputers to summarize and analyze outcome based education. Unpublished manuscript. Logan, UT: Utah State University.

Glossary of Terms

Aim Star: The point plotted at the intersection of the desired level of performance and the target date for achieving that level.

Alogrithm: A step by step procedure for solving an arithmetic problem.

Achievement test: An instrument used to assess the amount or level of learning that a student has acquired.

Arithmetic operation: A solution to the computation of a problem (e.g., addition, subtraction, multiplication, etc.).

Artificial intelligence: A branch of computer science devoted to researching ways to make the computer emulate those attributes we classify as human intelligence.

Automaticity: The immediacy of a student response to a stimulus (e.g., an arithmetic problem).

Bias: Unfairness in testing which can include interpretation of scores, test content, the development process and the procedures with which the test is administered.

CAMS (Chronometric Analysis of Math Strategies): An interactive computer program designed for classroom assessment of fluency in mathematics facts.

Ceiling: The point in test administration where the student receives no credit for all subsequent/more difficult items.

Chronometric analysis: An analysis of a student's response latency data which provides a profile of the student's strategy for solving the problems.

Copy: To imitate or reproduce an original item or model.

Criterion-referenced test: An assessment device designed to measure the student's ability to reach a designated level of performance on a specific task.

Curriculum-based measurement: An assessment system within which testing materials are derived from curriculum content, measurement is ongoing, and assessment information is used to develop instructional programs.

DEBUGGY: One of the earliest expert systems designed for education used for diagnosing student errors or "bugs" in the domain of place-value subtraction.

Diagnostic assessment: Formal and informal tests used to compare achievement levels, to sort people into groups, to identify skill deficits and to measure growth.

Diagnostic test: A measuring device used to determine causes of learning problems and/or specific strengths or weaknesses.

Error analysis: The individual analysis of a work sample to identify specific error types and set priorities for teaching.

Expert system: Computer software that emulates the functioning of a human expert in a particular field of knowledge.

Faulty algorithms: A skill or series of skills a student incorrectly has developed to solve specific types of problems.

Formative test: An assignment/test administered while learning is in progress.

Individualized education plan (IEP): A written account of educational objectives, strategies, curriculum modifications, and classroom accomodations for a child with learning and/or behavior problems; required by PL 94-142.

Interactive assessment: Software programs in which the computer acts in the role of examiner carrying out and analyzing the assessment data.

Item analysis: The process of collecting, summaryizing and using information about individual test items, especially information about pupils' responses to items.

Knowledge engineer: Person that develops a knowledge base and set of decision rules that represent the thinking process of a human expert or set of experts in a particular domain.

Line of progress: The trend of a student's progress drawn from the median of the three most recent days of data and the median of the three previous days of data.

Minimum 'celeration line: A line drawn from the student's current level of performance to the desired performance level (aim star) indicating the minimum level of acceleration or deceleration in the student's performance necessary to achieve the critrion level of performance.

Norm-referenced test: A valuative instrument in which a person's performance is judged in comparision with the average performance of others in a large reference group of similar age and/or educationallevels.

Summative test: A test administered at the end of a chapter or unit of instruction.

Test generation programs: Software programs designed to allow teachers to quickly and easily create an assessment instrument that covers specific skills and objectives at a given level of difficulty.

Test scoring and reporting programs: Software programs designed to score and summarize student responses on an assessment instrument.